Blockchain Architecture Design (KIT-061)

UNIT-I

What is blockchain?

A decentralized that enables multiple authoritative domains, who do not trust each other, to cooperate, coordinate and collaborate in a rational decision making process

Formal Definition of a Blockchain

- A Blockchain Is "an open, that can record transactions between two parties efficiently and in a verifiable and way"
- The keywords: Open (accessible to all), Distributed or Decentralized (no single party control), efficient (fast and scalable), verifiable (everyone can check the validity of information), permanent (the information is persistent)

The blockchain is a **decentralized distributed database** of immutable records. The technology was discovered with the invention of Bitcoins (the first cryptocurrency). It's a trusted approach and there are a lot of companies in the present scenario which are using it. As everything is secure, and because it's an open source approach, it can easily be trusted in the long run.

Bitcoin Blockchain and Ethereum Blockchain		
Topics	**Bitcoin**	**Ethereum**
Concept	Digital Currency	Smart Contracts
Founder	Satoshi Nakamoto	Vitalik Buterin
Release Method	Genesis Block Mined	Presale
Cryptocurrency Used	Bitcoin(Satoshi)	Ether
Algorithm	SHA-256	Ethash
Blocks Time	10 Minutes	12-14 Seconds
Scalable	Not yet	Yes

Features of Blockchain

Below are the most important features of Blockchain technology that has made it a revolutionary technology:

- SHA256 Hash Function
- Public Key Cryptography
- Distributed Ledger & Peer to Peer Network
- Proof of Work
- Incentives for Validation

Different types of blockchain:

- Public
- Private
- Hybrid/Consortium

1. **Public Blockchain**
    - Public Blockchain is publicly accessible and has no restriction on who can participate or be a Validator.
    - No one has complete control over the network.
    - This ensures data security and helps immutability because a single person cannot manipulate the Blockchain.
    - The authority on the Blockchain is equally divided among each node in the network, and due to this, Public Blockchains are known to be fully distributed.
    - Public Blockchains are mainly used for cryptocurrencies like **Bitcoin**, **Ethereum**, and **Litecoin**.
2. **Private Blockchain**

- A Private Blockchain (also know as Permissioned Blockchain) has restrictions on who can access it and participate in transaction and validation. Only pre-chosen entities have permissions to access the Blockchain. These entities are chosen by the respective authority and are given permission by the Blockchain developers while building the Blockchain application. Suppose there is a need to give permissions to new users or revoke permissions from an existing user, the Network Administrator can take care of it.
- Private Blockchains are mainly used in private organizations to store sensitive information that should be available only to certain people in the organization. Because Private Blockchain is a **Closed** Blockchain, the data is within the organization and out of reach from any external entities.

3. Consortium Blockchain

- In Consortium Blockchain, some nodes control the consensus process, and some other nodes may be allowed to participate in the transactions. Consortium Blockchain is like a hybrid of Public and Private Blockchain.
- It is public because the Blockchain is being shared by different nodes, and it is private because the nodes that can access the Blockchain is restricted. Hence, it is partly public and partly private.

The following table provides a detailed comparison among these three blockchain systems:

Property	Public blockchain	Consortium blockchain	Private blockchain
Consensus determination	All miners	Selected set of nodes	Within one organization

Property	Public blockchain	Consortium blockchain	Private blockchain
Read permission	Public	Public or restricted	Public or restricted
Immutability level	Almost impossible to tamper	Could be tampered	Could be tampered
Efficiency (use of resources)	Low	High	High
Centralization	No	Partial	Yes
Consensus process	Permissionless	Needs permission	Needs permission

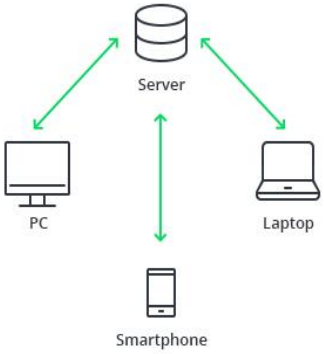
Design Primitives of Blockchain

- **Protocols for commitment:** Ensure that every valid transaction from the clients are committed and included in that blockchain within a finite time.
- **Consensus:** Ensure that the local copies are consistent and updated
- **Security:** Data needs to be tamper proof.
- **Privacy and Authenticity:** Data belong to various clients; privacy and authenticity needs to be ensured.

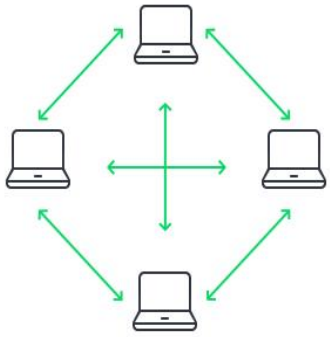
Difference between Blockchain database and Traditional databases

Properties	Blockchain	Traditional Database
Operations	Only Insert Operations	Can perform C.R.U.D. operations
Replication	Full Replication of block on every peer	Master Slave Multi-Master
Consensus	Majority of peers agree on the outcome of transactions	Distributed Transactions (2 phase commit)
Invariants	Anybody can validate transactions across the network	Integrity Constraints

Database vs. Blockchain Architecture



Client-server



P2P network

The traditional architecture of the World Wide Web uses a client-server network. In this case, the server keeps all the required information in one place so that it is easy to update, due to the server being a centralized database controlled by a number of administrators with permissions.

In the case of the distributed network of blockchain architecture, each participant within the network maintains, approves, and updates new entries. The system is controlled not only by separate individuals, but by everyone within the blockchain network. Each member ensures that all records and procedures are in order, which results in data validity and security. Thus, parties that do not necessarily trust each other are able to reach a common consensus.

To summarize things, the blockchain is a decentralized, distributed ledger (public or private) of different kinds of transactions arranged into a P2P network. This network consists of many computers, but in a way that the data cannot be altered without the consensus of the whole network (each separate computer).

Properties of Blockchain

There are four key features of blockchain:

- Decentralized Systems

- Distributed ledger
- Safer & Secure Ecosystem
- Mining

Double Spending Problem

It's a condition when one digital token is spent multiple times because the token generally consists of a digital file that can easily be cloned. It simply leads to inflation and organizations must bear a huge loss. One of the primary aims of Blockchain technology is to eliminate this approach up to the possible extent.

Is it possible to double spend in a Blockchain system?
Blockchain prevents double spending by confirming a transaction by multiple parties before the actual transaction is written to the ledger. It's no exaggeration to say that the entirety of bitcoin's system of Blockchain, mining, proof of work, difficulty etc, exist to produce this history of transactions that is computationally impractical to modify.

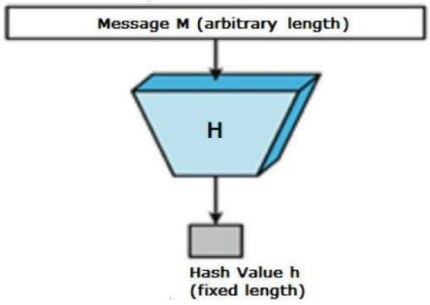
Hash function:

- Hash functions are extremely useful and appear in almost all information security applications.
- A hash function is a mathematical function that converts a numerical input value into another compressed numerical value. The input to the hash function is of arbitrary length but output is always of fixed length.
- Values returned by a hash function are called message digest or simply hash values. The following picture illustrated hash function.
- Hashing functions have a few properties that make them desirable for creating proof of work, a key concept in the Bitcoin network specifically:
- Hash functions turn an arbitrarily-large piece of data into a fixed-length hash output
- They are one-to-one: the same input will always provide the same hash output
- They are one-way functions: it's impossible to "work backwards", and reconstruct the input given a hash output.

The hash algorithm has certain unique properties:

1. It produces a unique output (or hash).
2. It is a one-way function.

In the context of cryptocurrencies like Bitcoin, the blockchain uses this cryptographic hash function's properties in its consensus mechanism. A cryptographic hash is a digest or digital fingerprints of a certain amount of data. In cryptographic hash functions, the transactions are taken as an input and run through a hashing algorithm which gives an output of a fixed size.



Properties of Hash Functions

In order to be an effective cryptographic tool, the hash function is desired to possess following properties −

- **Pre-Image Resistance**
    - This property means that it should be computationally hard to reverse a hash function.
    - In other words, if a hash function h produced a hash value z, then it should be a difficult process to find any input value x that hashes to z.
    - This property protects against an attacker who only has a hash value and is trying to find the input.
- **Second Pre-Image Resistance**
    - This property means given an input and its hash, it should be hard to find a different input with the same hash.
    - In other words, if a hash function h for an input x produces hash value h(x), then it should be difficult to find any other input value y such that h(y) = h(x).
    - This property of hash function protects against an attacker who has an input value and its hash, and wants to substitute different value as legitimate value in place of original input value.
- **Collision Resistance**
    - This property means it should be hard to find two different inputs of any length that result in the same hash. This property is also referred to as collision free hash function.
    - In other words, for a hash function h, it is hard to find any two different inputs x and y such that h(x) = h(y).
    - Since, hash function is compressing function with fixed hash length, it is impossible for a hash function not to have collisions. This property of collision free only confirms that these collisions should be hard to find.
    - This property makes it very difficult for an attacker to find two input values with the same hash.

- Also, if a hash function is collision-resistant **then it is second pre-image resistant.**

SHA256 Hash Function

The core hash algorithm used in blockchain technology is the SHA256. The purpose of using a hash is because the output is not 'encryption' i.e it cannot be decrypted back to the original text. It is a 'one-way' cryptographic function, and is a fixed size for any size of source text. To get a better understanding, let us look at an example below:

If you look at the first example, we are feeding the input as "Hello World" and getting an output as "a591a6d40bf420404a011733cfb7b190d62c65bf0bcda32b57b277d9ad9f146e". However, by just adding an "!" at the end, the output completely changes to "7f83b1657ff1fc53b92dc18148a1d65dfc2d4b1fa3d677284addd200126d9069". If we change "H" to "h" and "W" to "w", then the output value changes to "7509e5bda0c762d2bac7f90d758b5b2263fa01ccbc542ab5e3df163be08e6ca9".

- The complexity of this algorithm is as even the slightest change in the input can cause a massive change in the output. (**Avalanche Effect)**

Public Key Cryptography (Asymmetric Key Cryptography:

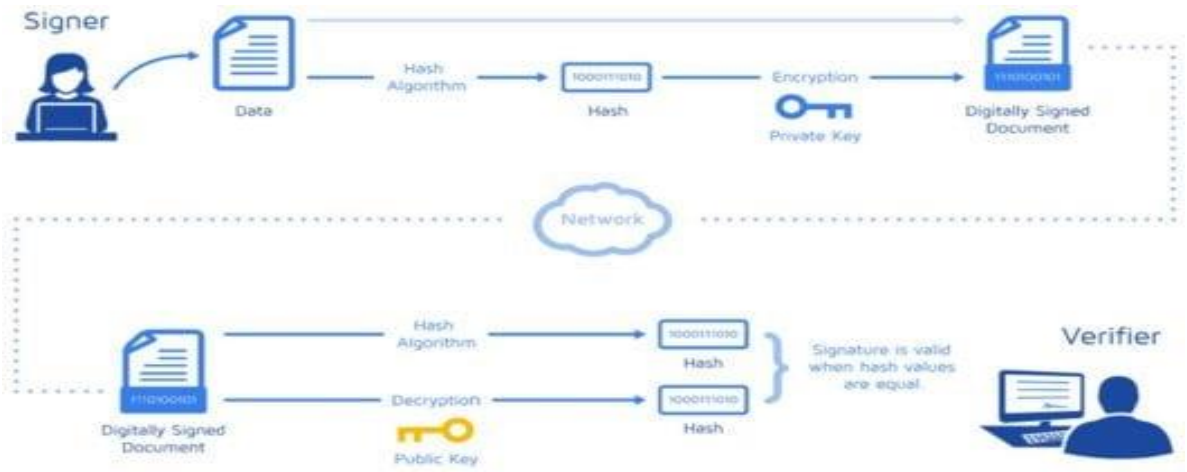
Two keys are used: Public key (Encryption) and Private key (decryption) for data encryption.

Digital Signature: (ensure the identity of user and prevent the non-repudiation attack)

This cryptographic technique helps the user by creating a set of keys referred as Public key and Private key. Here the Public key is shared with others whereas the Private key is kept as a secret by the user. To understand the roles of these keys, Let us look at the example below to get a better understanding:

If Chandler sends some bitcoins to Joey, that transaction will have three pieces of information:

- Joey's bitcoin address.(Joey's Public key)
- The amount of bitcoins that Chandler is sending to Joey.
- Chandler's bitcoin address.(Chandler's Public key)



Now all this data along with an encrypted digital signature is sent through the network for verification. The Digital signature is again a hash value achieved by the combination of the Chandler's bitcoin address and the amount he is sending to joey. This digital signature is encrypted by the private key. Once this data is received by a miner who has to verify this transaction, there are 2 process he does simultaneously:

1. He takes all the un-encrypted data like transaction amount and public keys of both Joey and Chandler, and feeds it to a hash algorithm to get a hash value which we shall call Hash1
2. He takes the digital signature and decrypts it using chandler's public key to get a hash value which we will call as Hash2

If both Hash1 and Hash2 are the same then it means that this a valid transaction

How Crypto currency Mining Works?

- Public and enterprise Blockchain are secure by design and are the continuously growing ledger on which all decentralized crypto currency and applications are built.
- Miners use high end computers to solve mathematical equations to verify transactions on the blockchain.
- Mining computers collect hundreds of pending transactions — also called a block — and turn them into a mathematical puzzle. The miner who finds the solution first gets rewarded.

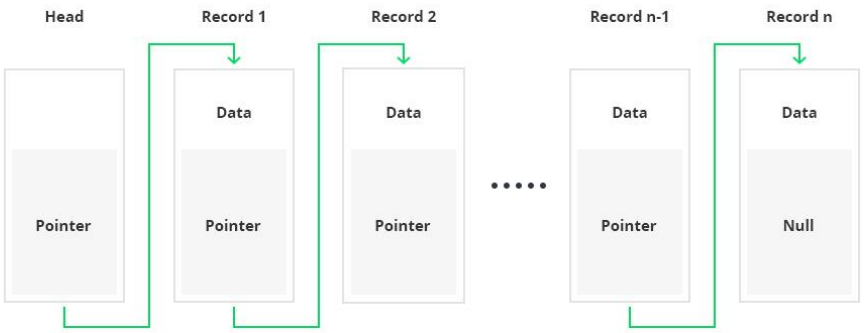
Blockchain Architecture

- Logically, a blockchain is a chain of blocks which contain specific information (database), but in a secure and genuine way that is grouped together in a network (peer-to-peer). In other words, blockchain is a combination of computers linked to each other instead of a central server, meaning that the whole network is decentralized.
- To make it even simpler, the blockchain concept can be compared to work done with Google Docs. You may recall the days of tossing over doc. documents and waiting for other participants to make necessary edits. These days, with the help of Google Docs, it is possible to work on the same document simultaneously.

- The blockchain technique allows digital information to be distributed, rather than copied. This distributed ledger provides transparency, trust, and data security.
- Blockchain architecture is being used very broadly in the financial industry. However, these days, this technology is employed not only for cryptocurrencies, but also for record keeping, digital notary, and smart contracts.

The structure of blockchain technology is represented by a list of blocks with transactions in a particular order. These lists can be stored as a flat file (txt. format) or in the form of a simple database. Two vital data structures used in blockchain include:

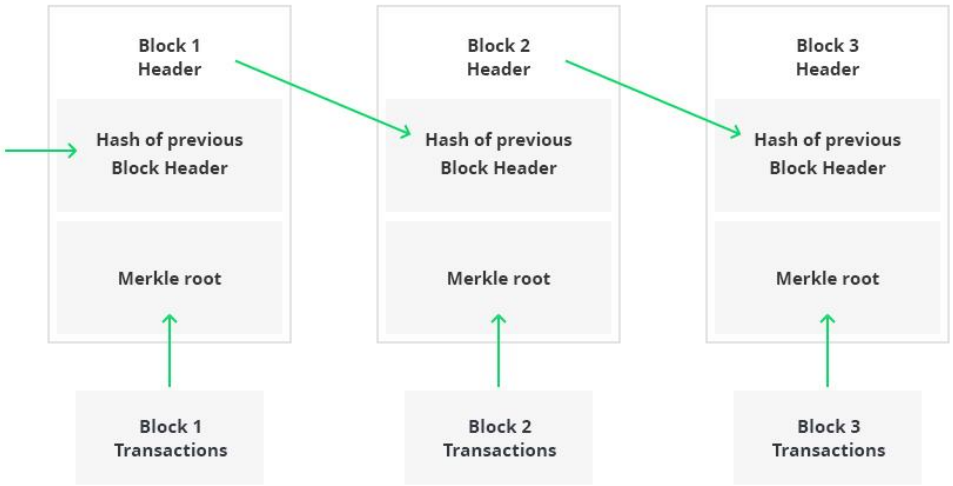
- **Pointers** - variables that keep information about the location of another variable. Specifically, this is pointing to the position of another variable.
- **Linked lists** - a sequence of blocks where each block has specific data and links to the following block with the help of a pointer.



Blockchain Hashing

Logically, the first block does not contain the pointer since this one is the first in a chain. At the same time, there is potentially going to be a final block within the blockchain database that has a pointer with no value.

Basically, the following blockchain sequence diagram is a connected list of records:



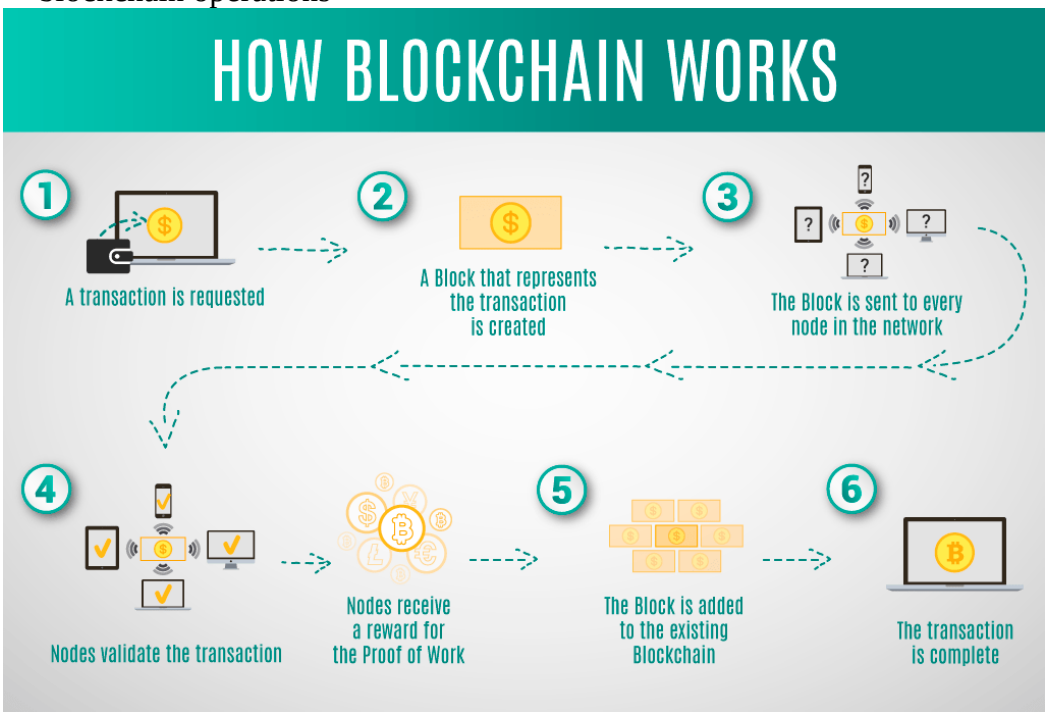
Blockchain architecture can serve the following purposes for organizations and enterprises:

- **Cost reduction** - lots of money is spent on sustaining centrally held databases (e.g. banks, governmental institutions) by keeping data current secure from cyber crimes and other corrupt intentions.
- **History of data** - within a blockchain structure, it is possible to check the history of any transaction at any moment in time. This is a ever-growing archive, while a centralized database is more of a snapshot of information at a specific point.
- **Data validity & security** - once entered, the data is hard to tamper with due to the blockchain's nature. It takes time to proceed with record validation, since the process occurs in each independent network rather than via compound processing power. This means that the system sacrifices performance speed, but instead guarantees high data security and validity.

Core Components of Blockchain Architecture: How Does It Work

These are the core blockchain architecture components:

- **Node** - user or computer within the blockchain architecture (each has an independent copy of the whole blockchain ledger)
- **Transaction** - smallest building block of a blockchain system (records, information, etc.) that serves as the purpose of blockchain
- **Block** - a data structure used for keeping a set of transactions which is distributed to all nodes in the network
- **Chain** - a sequence of blocks in a specific order
- **Miners** - specific nodes which perform the block verification process before adding anything to the blockchain structure
- **Consensus (consensus protocol)** - a set of rules and arrangements to carry out blockchain operations



Any new record or transaction within the blockchain implies the building of a new block. Each record is then proven and digitally signed to ensure its genuineness. Before this block is added to the network, it should be verified by the majority of nodes in the system.
The above blockchain architecture diagram that shows how this actually works in the form of a digital wallet.

Components of Blocks

Blocks are data structures whose purpose is to bundle sets of transactions and be distributed to all nodes in the network. Blocks are created by miners (discussed in more detail below).

- Blocks contain a block header, which is the metadata that helps verify the validity of a block.

- Typical block metadata contains:

    - **version** - the current version of the block structure

    - **previous block header hash** - the reference this block's parent block

    - **merkle root hash** - a cryptographic hash of all of the transactions included in this block

    - **time** - the time that this block was created

    - **nBits** - the current difficulty that was used to create this block

nonce ("number used once") - a random value that the creator of a block is allowed to manipulate however they so choose

These 6 fields constitute the block header. The rest of a block contains transactions that the miner has chosen to include in the block that they created.

Users create transactions and submit them to the network, where they sit in a pool waiting to be included in a block.

- A hash is like a fingerprint (long record consisting of some digits and letters). Each block hash is generated with the help of a cryptographic hash algorithm (SHA 256). Consequently, this helps to identify each block in a blockchain structure easily. The moment a block is created, it automatically attaches a hash, while any changes made in a block affect the change of a hash too. Simply stated, hashes help to detect any changes in blocks.
- The final element within the block is the hash from a previous block. This creates a chain of blocks and is the main element behind blockchain architecture's security. As an example, block 45 points to block 46. The very first block in a chain is a bit special - all confirmed and validated blocks are derived from the genesis block.
- Any corrupt attempts provoke the blocks to change. All the following blocks then carry incorrect information and render the whole blockchain system invalid.

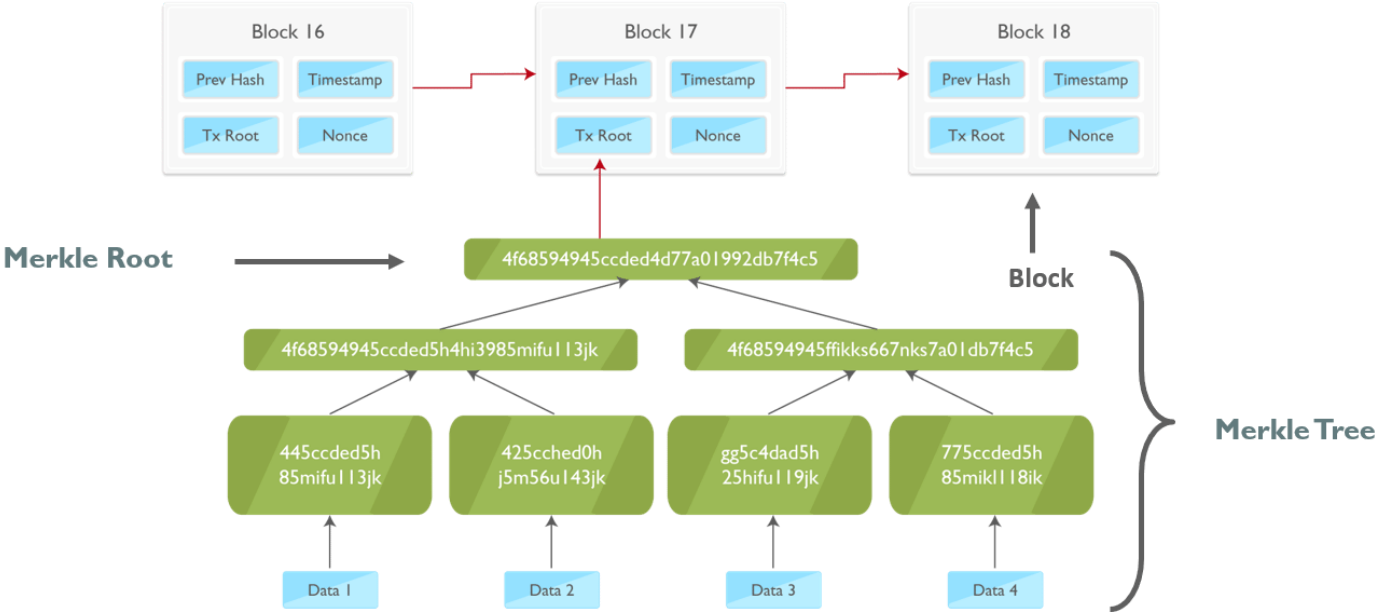
- On the other hand, in theory, it could be possible to adjust all the blocks with the help of strong computer processors. However, there is a solution that eliminates this possibility called **proof-of-work**. This allows a user to slow down the process of creation of new blocks.
- In Bitcoin blockchain architecture, it takes around 10 minutes to determine the necessary proof-of-work and add a new block to the chain. This work is done by miners - special nodes within the Bitcoin blockchain structure. Miners get to keep the transaction fees from the block that they verified as a reward.
- Each new **user (node)** joining the peer-to-peer network of blockchain receives a full copy of the system. Once a new block is created, it is sent to each node within the blockchain system. Then, each node verifies the block and checks whether the information stated there is correct. If everything is alright, the block is added to the local blockchain in each node.
- All the nodes inside a blockchain architecture create a **consensus protocol**. A consensus system is a set of network rules, and if everyone abides by them, they become self-enforced inside the blockchain.

For example, the Bitcoin blockchain has a consensus rule stating that a transaction amount must be cut in half after every 200,000 blocks. This means that if a block produces a verification reward of 10 BTC, this value must be halved after every 200,000 blocks.
As well, there can only be 4 million BTC left to be mined, since there is a maximum of 21 million BTC laid down in the Bitcoin blockchain system by the protocol. Once the miners unlock this many, the supply of Bitcoins ends unless the protocol is changed.

- To recap, this makes blockchain technology immutable and cryptographically secure by eliminating any third-parties. It is impossible to tamper with the blockchain system; as it would be necessary to tamper with all of its blocks, recalculate the proof-of-work for each block, and also control more than 50% of all the nodes in a peer-to-peer network.

Merkle Tree roots

Merkle Tree also known as 'hash tree' is a data structure in cryptography in which each leaf node is a hash of a block of data, and each non-leaf node is a hash of its child nodes.



The benefit of using the Merkle Tree in blockchain is that instead of downloading every transaction and every block, a “light client” can only download the chain of block headers. Also, if someone needs to verify the existence of a specific transaction in a block, then he doesn’t have to download the entire block. Downloading a set of a branch of this tree which contains this transaction is enough. We check the hashes which are just going up the branch (relevant to my transaction). If these hashes check out good, then we know that this particular transaction exist in this block.

Mining process

- ”Mining is the mechanism that allows the blockchain to be created securely and in a decentralized manner. It provides the basis for the cryptocurrency system and enables a peer-to-peer network without a central authority.

- Miners validate new transactions and record them on the global ledger (blockchain). On average, a block (the structure containing transations) is *mined* every 10 minutes.
- Miners compete to solve a difficult mathematical problem based on a cryptographic hash algorithm. The solution found is called the *Proof-Of-Work*. This proof proves that a miner did spend a lot of time and resources to solve the problem. When a block is 'solved', the transactions contained are considered *confirmed*, and the bitcoin concerned in the transactions can be spend.